

Applied Linear Regression, 22S:152

Instructor: Cowles

Lab 1

Sep. 4, 2002

1 If you want SAS for your own computer

SAS for Students, Windows version, is available at the IMU Bookstore at a reduced cost for students using the software for academic purposes. This license needs to be renewed in October. See the SAS Information page,

<http://www.its.uiowa.edu/its/cs/software/sas.html>
for product information.

2 Getting started in the ITC

ITC computers will display a login screen. Users should enter their HawkID and password in the spots provided. Students can find their HawkID and default passwords on the ISIS system.

3 Downloading files from the course web page

Double click on Netscape. Enter the address of the my web page in the location box at the top of the screen.

```
www.stat.uiowa.edu/~kcowles
```

Then click on “Course homepages” and “22S:152”

Click on “Datasets,” and when the next screen appears, click on the underlined link “Datasets.” Three types of files may be accessed:

- Files ending in .dat are data files for your use with the software package SAS.
- Files ending in .info contain descriptive information about datasets.
- Files ending in .txt are datasets for use in a different class with a different software package.

To download a data file for use in this lab:

- Right click the file name (use the *right*, not left, mouse button)
- In the dialog box that opens, left click “Save link as” (use the *left* mouse button)
- In the “save as” dialog box move to the “temp” directory on the C: drive and click “Save.”

- (If you preferred to save the file on your own disk in drive A: so that you could use it on a different computer later, you would move to drive A: in the dialog box before saving.)

Left click the “OECD.info” file to read a description of the “OECD” dataset. (Note that the system on which I stored these files alphabetizes upper case letters before lower case letters.) Click “Back” to get back to the list of datasets. Then download the file “OECD.dat” according to the directions above.

4 Other useful features on the course web page

Return to the main course web page and click “Web resources.” Note that there is a directory of the locations and hours of all the campus ITCs, as well as links to other useful sites.

Lecture notes, homework assignments, and lab handouts are posted under “Handouts.” These are in PDF format, which may be read and printed in the ITCs and on most other computers. (Acrobat Reader, the software that reads and prints PDF files, can be downloaded for free.)

5 Accessing SAS

Double click on SAS 8.2e.

You will get a screen that shows:

- a menu bar
- a log window
- a program editor window

6 Entering commands and programs

Click in the program editor window. You may now type commands and programs in this window.

7 How SAS programs and commands are organized

Use a *DATA step* to organize your data by creating a SAS dataset. Then use *PROC steps* to analyze your data using SAS procedures. Once you have created a SAS dataset, you may apply any SAS procedures to it without recreating the dataset.

DATA and PROC steps consist of SAS *statements*. Each statement must end with a semicolon. Most statements include one or more *keywords* that must be spelled exactly as shown.

8 The DATA step: Creating a SAS dataset

Before it can process data, SAS must read in the data in the form of a table with

- a row for each *observation*
- a column for each *variable*

You must choose a name for the entire dataset and a name for each variable. SAS has the following rules for names:

- SAS names must begin with a letter or an underscore. The remaining characters in a SAS name can be letters, numbers, or underscores. There must be no embedded blanks.

SAS distinguishes between two types of variables:

- *numeric variables*, which contain only digits, decimal points, and minus signs, and with which arithmetic operations may be done
- *character variables* (all other kinds of data)

9 Reading data in from an existing datafile

You have saved the file “OECD.dat” in the “temp” directory. Use an “infile” statement to tell SAS to use it.

```
data OECD ;                * gives dataset a name for SAS ;
infile 'c:\temp\OECD.dat' ; * tells SAS where the data is ;
input name $13. pcgdp pch beds los docs infmort ;
                        * $13 after name identifies character vbl
                        * and tells SAS maximum length of values ;
run ;                      * end of data step ;
```

Type these lines into the program editor window. (You do not have to type the asterisks and comments after them.) To make SAS run these statements and create the dataset, use the mouse to highlight the block of statements and then click on the icon of the running man.

SAS will use the log window to tell you what it has done. Be sure to check the log window for any error messages. If any errors are reported, click in the program editor window to make it active. Correct the errors in the code and then rerun the block of code.

Now we can apply any SAS procedures to this dataset, either by typing in additional code in the “Program Editor” window or by using one of the menu-driven tools.

10 Displaying the dataset

First we will use `proc print` to list the entire dataset on the screen. Displaying all or part of the dataset is always wise in order to make sure that you have read the data in correctly.

```
proc print data = OECD ;    * word after "data = " must be the name of
                          * a SAS dataset created in a data step ;
run ;                      * end of print procedure ;
```

You can restrict how many records are displayed as follows:

```
proc print data = OECD (obs = 10) ;
run ;
```

You can also use the `var` statement to restrict which variables are displayed:

```
proc print data = OECD ;
var name pcgdp infmort ;   * variable names must be from the data step in
                          * which the dataset was created or modified;
run ;
```

11 Proc univariate for summary statistics of quantitative variables

`Proc univariate` is a quick and convenient way to get a complete numeric summary of the distribution of a quantitative variable. You can also request plots (which are ugly but easy to print).

```
proc univariate plot data = OECD ; * leave out word "plot" if you dont't
                                  * want or need the plots ;
var pcgdp infmort ;
run ;
```

12 Interactive data description with “Insight”

We will use “Insight” to get additional descriptive statistics and plots regarding these data.

From the menu bar, select the following sequence of choices:

- Solutions
 - Analysis
 - * Interactive data analysis

In the window that appears, you must specify which dataset you wish to use. Do so by clicking

- Library: Work

- Dataset: OECD
- * Open

Use the “Analyze” Menu to do some data screening on the following continuous variables:

- infmort (infant mortality rate)
- pcgdp (per capita gross national product)

After you have selected either “Histogram,” “Boxplot,” or “Distribution,” you will be prompted to select a variable. Click the one you want; then click on “Y.”

To get both a graphical and a tabular summary of the variable, use “Distribution.” After selecting a variable, click “Output.” Click on the choices of output that you want. To get a histogram with a normal curve superimposed, choose “Density estimation” and check “Normal” under “Parametric estimation.”

Are there any implausible values?

Are the distributions skewed or symmetric?

13 Creating new variables

Choose one of the variables and experiment with various transformations from Tukey’s ladder of transformations to see which will approximately symmetrize the variable. You may do this in one of two ways :

- Use “Edit” “Variables” from the Insight menu.
- Exit out of Insight and return to the Program Editor by clicking in the window showing the data in spreadsheet form. Then pull down the “File” menu and choose “End.” Then edit your data step in the “Program Editor” window to create the new variables. For example, you could insert new lines:

```
sqrtinfmt = sqrt(infmort) ;
loginfmt = log(infmort) ;      * natural log (log 10 is log10 function) ;
invinfmt = - 1.0 / infmort ;
```

in between the “input” statement and the “run” statement. Statements such as this that create new variables must be placed after the “input” statement in which the original variables are read and inside a data step.

Then resubmit the whole datastep (by highlighting the code and clicking the running-man icon). Then reinvoke Insight using the menu bar.

14 Printing and Saving Files

The easiest way to print SAS code and output from procedures run from the program editor window is to copy and paste the output from the SAS window into Microsoft Word; then edit and print from Word. Output from Insight can be printed directly from Insight using the File menu.

To save a file from SAS, click in the window whose contents you want to save. Go to the file menu and choose “Save as”. If you want to save your SAS program and work with it again later, you must save it on your disk in the A: drive. SAS will automatically give the file extension “.sas” to SAS commands and programs. For example, to name a SAS program “myprog,” you would type

a:myprog

in the box for the name of the file. You can read it back into SAS at another time using the “File” menu.

15 When you have finished...

Be sure to exit from SAS using the File menu, and to log out of the computer using the icon on the desktop.